On the Non-Linear Relation between Trade and Distance: New Border-Effect Estimates Based on Region-to-Region National and International Flows

Nuria Gallego* (nuria.gallego@uam.es)

Carlos Llano* (carlos.llano@uam.es)

* Departamento de Análisis Económico: Teoría Económica e Historia Económica. Facultad de Ciencias Económicas y Empresariales. Universidad Autónoma de Madrid. Campus Cantoblanco. 28049 MADRID.

JEL Codes: F14, F15
Keywords: border effect, distance, interregional trade, region-to-region international trade, European integration.

Abstract

The existence of large border effects is one of the main puzzles of international macroeconomics. The seminal paper by McCallum found that trade between any two Canadian provinces was (on average) 22 times greater than trade between any Canadian province and any US. state. Although various authors have estimated internal and external border effects for the whole EU and some specific European countries, none has done so in the manner that seminal paper, stymied by lack of data on region-to-region international trade flows. This paper uses a novel dataset that captures intra- and inter-national truck shipments between Spanish regions and regions in eight European countries. It computes internal and external border effects, offering novel results for aggregate flows and the importing countries, and estimates several specifications of the gravity equation, so as to tackle such issues as the multilateral resistance term, heteroskedasticity, and zero flows and non-linear relation between trade and distance. For non-linear relations, we use a new strategy based on segmented distance, thereby achieving the disappearance of the internal border and obtaining a persistent external border of 7.

1. Introduction

The existence of large border effects is one of the main puzzles of international macroeconomics (Obstfeld and Rogoff, 2000). The seminal paper by McCallum (1995) found that trade between any two Canadian provinces was (on average) 22 times greater than trade between any Canadian province and any US. state. Since then, many authors have repeated the exercise with other countries[1] and other spatial units, whether countries, regions, provinces or even zip codes (see Table 1 for summary).

For the European Union (EU), certain papers have estimated the relevance of international borders by comparing a European country's domestic trade volume with its international trade volume (Head and Mayer, 2000; Minondo, 2007; Chen, 2004). Similar analyses have been produced at the sub-national level so as to compute *external border effects*. These have taken a country's regions (or provinces) as their point of reference and counted how many more times they traded with the rest of the country (as a whole) than with other countries (Gil et al., 2005; Ghemawat et al., 2010).

In parallel, we also find estimates of *internal border effects*, defined as how much more trade some region (province) of a given country conducts with itself than with any other region (province) of the same country. Wolf (1997, 2000), for example, while investigating market fragmentation in the United States, found intra-state trade unduly high in relation to inter-state trade. Later, Hillberry and Hummels (2008) analysed the impact of geographical frictions on trade, using truck deliveries within United States at different spatial levels. They found that internal border effects would disappear in the US. as the spatial units became very fine. Similarly, Combes et al. (2005) and Garmendia et al. (2012), taking into account social and business networks, investigated the narrowing of internal border effects at the province level (Nuts 3) for France and Spain, respectively.

To the best of our knowledge, no one has yet produced a pristine estimate of border effects in Europe. No one, that is, has estimated how much more trade a particular region of a European country conducts with another region of the same country than with a third region

[1] Japan (Okubo, 2004), United States (Wolf, 2000; Hillberry, 2002; Hillberry and Hummels, 2003; 2008; Millimet and Osang, 2007), the European Union (Chen, 2004; Nitsch, 2000, 2002; Evans, 2003), Germany (Shultze and Wolf, 2008), Russia (Djankov and Freund, 2000) and Brazil (Daumal and Zignago, 2008), among others.

in another European country. The reason is lack of data on region-to-region trade flows between Europe's countries. Thus the most ambitious attempts to measure the effect of international borders on inter-regional trade structures in the European Single Market are either indirect or restricted to border regions with intense bilateral trade relations (Lafourcade and Paluzie, 2011; Helble, 2007).

Moreover, whenever an *external border effect* has been computed—on the basis of flows between sub-national units on both sides of a national border (McCallum, 1995; Anderson and van Wincoop, 2003; Feenstra, 2002, 2004)—only inter-regional flows between contiguous countries (e.g., Canada-US.) have been considered. The actual effects of a national-border crossing have thus been mixed with those of high economic integration and cultural and historical similarities. It would therefore be most interesting to have an alternative estimate of Anderson and van Wincoop's border effect, one between the Canadian provinces and the regions of a non-contiguous country: the Mexican states, for example. Space being non-neutral, we should keep in mind that a Canadian province wishing to deliver a product by truck to regions in another country must either trade with the US. or send the truck across it. The United States, on the other hand, can deliver products by truck to two contiguous countries, Canada and Mexico. Similarly, Spain can trade with three contiguous countries (Portugal, France and Andorra), and with many others once its trucks have crossed France. Since competing destinations so thoroughly condition international trade (Anderson and van Wincoop, 2003), and different European countries share such different levels of economic and cultural integration, it would be interesting to compute external border effects *à la* McCallum—but to do so for region-to-region trade between non-contiguous European countries as well. This would give us a first insight into the true roles played by various national borders.

In this paper, using a unique dataset that captures region-to-region intra- and inter-national trade, we estimate *internal* and *external border effects*, contrasting the intra- (Nuts 2) and inter-regional trade between Spain's regions against the inter-regional trade between Spanish regions and those of eight other European countries. To do so, we test several specifications of the gravity model (McCallum 1995; Feenstra 2002, 2004; Anderson and van Wincoop 2003; Gil et al, 2005), producing benchmarks for the results of our novel dataset.

We then dig deeper into one of the most recent proposed solutions to the border-effect puzzle. The literature advances alternative factors to explain the greater intensity of trade within regions and countries[2]. Some recent papers describe the border effect as a "spatial aggregation artefact" (Hillberry and Hummels, 2008; Llano-Verduras et al., 2011) or as a mismeasurement of the distance variable (Head and Mayer, 2000, 2002). We aim to shed light on the non-linear relationship between distance and trade. Like the aforementioned analysts, we start by incorporating a quadratic distance term, so as to capture the fast decrease in trade flows over the shortest distances. Instead of varying the size of the exporting unit, however, we focus on the spatial disaggregation of the destination (from country to region). Then, to deal with this non-linearity, we suggest an alternative strategy, which considers alternative sub-divisions of the sample by distance travelled. Our approach can thus address the *fractal*[3] non-linear relationship between trade flows and distance (Brakman et al., 2009). In our case, the fractal dimension of the non-linearity appears with changes in spatial level (country, region, province) and when flows cross borders of differing nature or thickness (own region; inter-regional areas within a country; inter-regional areas between countries with different levels of integration).

In sum, we believe this paper contributes to the previous literature in the following ways: (1) It produces the first estimates for the *external border* of European regions by means of region-to-region flows, just as McCallum, Feenstra or Anderson and van Wincoop did for Canada and the US. These estimates confirm that trade integration is even higher between European regions than between North America's equivalent spatial units. (2) It computes *external* and *internal border effects* simultaneously for inter-regional international flows between one country (Spain) and its eight main European partners. Whereas previous papers considered only inter-regional trade between two contiguous countries, this approach sheds new light on the effect of different national borders. (3) Like other papers, ours obtains border effects that shrink along with the size of the exporting unit (from state to zip code in Hillberry and Hummels, 2008; from region to province within the exporting country (Spain) in Llano-Verduras et al., 2011). However, it obtains this decrease simply by dividing up the spatial unit of the importer (from country to regions). This result probably masks the

[2] Among them the presence of external barriers to trade (e.g., tariffs and non-tariff barriers), endogenous responses (agglomeration economies), information barriers (Rauch, 2001), social and business networks (Combes et al., 2005), home bias in the preferences and the heterogeneity of firms (Evans, 2003; Chaney, 2008) and the misspecification of econometric models (Anderson and van Wincoop, 2003).

[3] Term used in spatial economics to describe a phenomenon observed persistently at different spatial levels (i.e., spatial agglomeration of economic activity at the region, country or urban level).

mismeasurement of the *external border effect* in region-to-country data, a mismeasurement due to an inappropriate assumption about the distribution of trade within the importing country's regions. (4) Finally, it suggests an alternative strategy for tackling the non-linear relationship between trade and distance: namely, alternative divisions of the sample by distance travelled. This strategy produces interesting results: (i) a null *internal border effect* along with a persistent *external border effect* of factor 7; (ii) (perhaps more surprisingly) a variation in the elasticity of distance when the distance is segmented in three alternative ways: especially by the well-known power series known as the *Fibonacci sequence*.

The rest of the paper is organized as follows. Section 2 revisits the literature on border effects. Section 3 briefly summarizes our method for estimating a region-to-region trade dataset for the Spanish case and offers a descriptive analysis of new trade flows. Section 4 describes the alternative specifications of the gravity equation used in our analysis. Section 5 presents the results obtained with different specifications of the gravity equation. The final section summarizes our main conclusions.

2. The Border Effect

After McCallum's seminal contribution (1995), Helliwell (1996, 1997, 1998), Hillberry (1998) and Anderson and Smith (1999) confirmed the existence of an *external* border effect between Canada and the United States. They used province-to-state data and, of course, considered only one international border. In his estimate, Hillberry (1998) used data from the US. Transportation Department's commodity-transportation survey and obtained results similar to McCallum's and Helliwell's. Helliwell and Verdier (2001) quantified the *internal* and *external* border effects between Canadian provinces and US. states for 1991–1996. In their results, the *internal* border varied from 27 to 2. They also found an *external border* suggesting that one Canadian province had traded about fifteen times more often in 1991 (and ten more times in 1996) with other domestic subunits than with any US state.

Wei (1996) studied external border effects for OECD countries for 1982–1994. He found that these countries (on average) traded ten times more often within themselves than with other countries. With the application of a gravity equation "augmented" by additional variables (e.g., remoteness, common language, common frontier), the border effect dropped to a factor of 2.6. Helliwell (1997) used the same dataset as Wei (1996) for the year 1990.

He considered a more completed specification of border effects, with a more elaborate common-language variable and a different measure of remoteness, and found an external border effect roughly twelve times greater for internal trade in OECD countries than for their trade with other countries.

Some years later, Anderson and van Wincoop (2003) and Feenstra (2002, 2004) revolutionized the literature of border effects and the gravity equation. Digging deeper into the theoretical underpinnings of the gravity equation, they showed that border effects were being overestimated whenever the model specification did not control for the effects that non-observable price indices on each trading partner (multilateral resistance).

<< Table 1 about here >>

Nitsch (2000) measured the impact of national borders on international trade between EU. countries for 1979–1990, finding that intra-national trade was on average around ten times greater than inter-national trade. He then analyzed this effect's evolution over time in two samples: for 1979–1990 (Spain and Portugal excluded) and for 1983–1990 (Spain and Portugal included). His results showed that the border for the first sample dropped gradually from 9 to 7 and the border for the second from 12 to 11. Unlike previous authors (Wei, 1996; Wolf, 1997), Nitsch made his estimation with an alternative measure of intra-national distance: the square root of a country's area multiplied by a scaling constant. Similarly, Head and Mayer (2000) focused on the sources of Non-Tariff Barriers (NTB) to trade in the EU. For average flows in 1984–86, they found that European purchases tended to be sixteen times higher from the domestic country than from any other European country. For 1993–95 the same effect fell to a ratio of 13. Additionally, Head and Mayer (2002) showed how border-effect estimates were affected by distance measurement (mainly internal distance). They also found that border and adjacency effects in Europe had diminished over time but not vanished. Chen (2004) focused on estimating and explaining border effects between seven EU countries using flows specific to 78 manufacturing industry for the year 1996. She found that with controls in place for multilateral resistance the border effect decreased. Her obtained border effect suggested that intra-national trade was roughly six times greater than inter-national trade, *caeteris paribus* the other variables. Helble (2007), finally, estimated the border effect for France and Germany, using extensive data on trade between each region and fourteen EU. countries and combining inter-national trade and intra-national transport flows. According to his results, France traded roughly eight times more often with itself than with any other EU country; Germany, about three times more often.

For the border effect in Spain, Gil et al. (2005) examined the magnitude of the *external border effect* with bilateral trade flows between each of the seventeen Spanish regions (Nuts 2) and 27 OECD countries for 1995–1998. Using a panel-data technique with random effect, they found that, on average, exports from any Spanish region to the rest of Spain (as a whole) exceeded inter-national exports by a factor of 20 and inter-national imports by a factor of 24. All of these results were obtained with controls for size, distance, contiguity, island geography and membership in the EU or EFTA.

Thereafter, at least four papers, all using region-to-country trade flows, revised the Spanish border effect. Requena and Llano (2010) estimated the *internal* and *external border effect* at the regional level (Nuts 2) using industry-specific flows. Their dataset included intra- and inter-regional trade flows for each of the seventeen Spanish regions as well as each region's international trade flow with each of the twenty-eight OECD countries. The authors found that the *internal border effect* reached an average value of 17: i.e., the average Spanish region tended to trade seventeen times more with itself than with the rest of the country. According to the authors' *external border effect*, the average Spanish region tended to trade thirteen times more with the rest of Spain (as a whole) than with any other country in the sample. With a similar dataset, Ghemawat et al. (2010) focused on Cataluña's *external border effect*, contrasting trade between Cataluña and the rest of Spain (as a whole) against its trade with twenty-two other OECD countries. The results suggested a decline in Catalonia's *external border effect* for 1995–2005 from a factor of 80 to a factor of 29. However, when the same analysis was repeated for 2005 and exclusively for Catalonia's trade (exports + imports) with contiguous France, the external border effect decreased to 23. Llano-Verduras et al. (2011) revised the *external border effect* in Spain using flow data at two different spatial scales: namely, regions (Nuts 2) and provinces (Nuts 3). They found that the size of the border effect depended largely on the unit of spatial measurement. Unfortunately, this final paper—while it does vary the spatial scale for Spanish units: from regions (Nuts 2) to provinces (Nuts 3)—always scales the foreign partner at the country level. Garmendia et al. (2012) re-estimated Spain's *internal border effect* with province data (Nuts 3), taking into account social- and business-network effects.

In summary, to the best of our knowledge, no paper estimating the border effect in Europe, and particularly in Spain, has yet managed to replicate the interregional-trade approach used for Canada and the US. Moreover, all previous estimates of regional border effects in Europe have had to resort to comparisons between regions and countries (or "parts of countries"). Never has the same spatial unit been used on both sides of a national border. Furthermore, only a few papers have pondered the non-linear relation between trade and distance.

3. The Data

We should state at the outset that there is no official data on region-to-region international trade flows for any country in the EU. Gallego and Llano (2012), however, have laid out a method for estimating region-to-region international flows between Spain and eight European countries[4]. It combines region-to-region freight statistics for Spanish trucking with international price indices (deduced from official trade data[5]) for each region-country variety (cf. Appendix for a summary of the method).

This novel dataset for region-to-region international trade flows was connected with equivalent data on (intra- and inter-regional) trade flows within Spain. This second dataset was generated for the C-intereg project (www.c-intereg.es) and has been the object of previous analysis, as in some of our benchmarks (Garmendia et al., 2012; Llano-Verduras et al., 2011; Ghemawhat et al., 2010; Requena and Llano, 2010; Llano et al., 2009).

The result is a unique dataset on region-to-region flows for intra-regional, inter-regional and inter-national flows into and out of the regions of Spain (Nuts 2) and the regions of Spain's eight main European partners. Each flow can be described by equation [1]:

[4] Although for the sake simplicity we use the label EU, our sample of countries does not fall under any specific administrative category. Moreover, we consider certain countries, like Andorra, as single-region countries.

[5] For most of our EU countries, we use two main sources for the inter-national bilateral flows of goods: (1) Trade statistics on intra-EU trade, which register bilateral flows between pairs of countries, both in volume and in monetary units; for certain countries, like Spain, the trade data identify the exporting or the importing region but never both simultaneously. (2) Transport statistics on intra-national and inter-national freight flows, which in some cases (e.g., road freight) provide information on the type of product transported (quantity) as well as on the regional origin and destination of the flows. Our method aims to build up a region-to-region trade dataset by combining these two sources: (1) region-to-region flows in quantities (road-freight statistics) and (2) specific region-to-country trade prices (from the official trade statistics).

$$\hat{T}_{ij}^{eugR} = \left(F_{ij}^{eugR} \hat{P}_{ij}^{eugR} \right) \qquad [1]$$

Estimated trade flows $\hat{T}_{ij}^{eugR}$ of products g traveling by Spanish trucks from region i in country e to region j in country u were obtained by the combination of estimated trade prices $\hat{P}_{i.}^{eugR}$ and actual *freight flows* (in tons) F_{ij}^{eugR} as delivered by Spanish trucks[6]. Note that if e = *Spain*, equation [1] captures Spanish exports to regions in our sample of eight European countries (cf. list of countries and products in the Annex). In this case, only Spanish exports of goods transported by Spanish trucks are considered. Because of the characteristics of our road-flow dataset, we exclude flows where the regional partner is a Spanish island (I=15 for Spain).

Our distance variable is the *mean actual distance* covered by Spanish trucks between each pair of trading regions, as reported in the survey published by Spain's Ministry of Public Works and Transport (Ministerio de Fomento). This variable has the virtue of capturing the *real distance* travelled by trucks between actual points of departure and destinations. In this sense, it is superior to the variable used by other authors, where intra-national and/or international distance is either an a-priori estimate based on the great circle distance between main cities weighted by population or an ad-hoc estimate by mathematical approximation. By using actual distance, we should, in theory, be able to account for region-to-region inter-country links that are not attributable to the mere allocation of population. There are specific regional endowments or specificities that weighted distance tends to mask. This difference alone should set this paper apart from papers that use region-to-country flows and assume that the distance between the exporting region and the importing country (as a whole) is equivalent to the weighted average of the linear distance between the exporting region and the importing country's main cities (weighted by population). The Ministry's survey also includes the actual distance travelled by trucks for inter- and intra-regional deliveries. Crucially, this allows us to avoid choosing alternative ad-hoc intra-regional distances, which alter results on border effects (Head and Mayer, 2002). Because the actual distance travelled by each truck between each pair of regions can vary in every year, and because with the aim of eliminating the risks of endogeneity, we have arrived at our intra- and inter-regional

[6] Although trade and freight statistics offer data on trading volume (tonnage), we use a different notation for each: Q = trade-statistic volume, F = freight-statistic volume.

distances by averaging the distances observed in all deliveries from 2004 to 2007 for each specific dyad *i-j*. Regional GDPs for the EU regions under consideration are published by Eurostat.

3.1. Descriptive Analysis

Before proceeding to the econometric analysis, we will briefly analyze the novel dataset. In the Appendix the new dataset for international flows is compared with the official trade data, so as to detect and discard inconsistencies. Next, we will identify the main inter-regional aggregate flows together with Spanish inter-regional flows.

<< Figure 1 about here >>

Figure 1 plots the spatial concentration of exports delivered from three important Spanish regions, corrected by the product of the GDPs of the trading partners. It is worth mentioning the novelty of these three maps, which reveal the hitherto unknown region-to-region dimension of Spain's trade relations with eight European partners. It is important, also, to highlight that in all three cases intra-national trade flows are clearly greater than inter-national flows. The intensity of the color shows a clear discontinuity in the relevance of trade flows between Spanish and European markets, even when three of the country's main exporting regions are considered. As we will see in the next sections, this result leads to positive and significant *external border effects* for all exporting regions and all importing countries. Note that for all three regions the most intense interregional trade flows within Spain are concentrated in the closest regions. This will be tested by contiguity dummies.

In addition, like some recent authors (Garmendia et al., 2012; Llano-Verduras et al., 2011; Hillberry and Hummels, 2008), we offer here a first view of the distribution of trade (always region-to-region) as it depends on distance travelled by trucks, for both domestic and international deliveries. Like them, we also use a kernel regression to generate a nonparametric estimate of the relationship between distance and the intensity of Spanish regional export flows[7].

[7] We use the Gaussian kernel estimator in STATA, with n = 100 points and the estimator calculating optimal bandwidth.

<< Figure 2 about here >>

Figure 2 plots the distribution of domestic and international flows (exports) for each region against those for the rest of Spain's regions and the eight European countries. Note that trade flows are corrected by the GDP of each exporting/importing region. To illustrate the *fractal dimension of the non-linear relation between trade and distance*, the figure has a separate plot for the kernel regression of each kind of trade flow: i.e., intra-regional flows within Spain, inter-regional flows within Spain and inter-regional exports from Spanish regions to regions in the eight countries. To bring out the great differences in intensity, the graph displays two different scales: one for intra-regional flows (left axis) and one for the remaining flows (right axis). Moreover, to emphasize the similar shape of each kernel distribution, the "international flows" kernel is plotted twice: with its natural scale and re-scaled at a factor of "x7" (in line with the *external border effect* reported in Table 3). In distinguishing the great differences in the relative intensity of the flows, we can also see the *regularity* of the non-linear relationship between trade and distance over the shortest distance. By mixing together different types of flows, other papers have emphasized the sharp decrease on the intensity of trade over the shortest distances (e.g., 700 km). Our approach shows how length of flow varies by kind of flow.

From this analysis we can conclude that, regardless of flow type, the bulk of trade takes place over short distances and beyond a certain point the negative effect of distance falls off dramatically. Hence the relevance of territorial disaggregation. Llano-Verduras et al. (2011) and Garmendia et al. (2012) have shown that, with insufficient territorial disaggregation of trade, the gravity equation may lead to an overestimation of the border effect and an underestimation of the distance effect. As we will see in the next section, this overall effect can arise not just when regions are used instead of provinces, but also when countries are used instead of regions. Because of the *fractal dimension* of this non-linearity, moreover, sharp decreases in trade intensity may or may not coincide with the administrative units where the flows are allocated (and where the borders are!). It would thus be interesting to consider econometric procedures flexible enough to control for that.

4. The Empirical Model

As in most of the articles cited previously, the backbone of our investigation is the gravity equation, where the intensity of trade between any two locations (regions or countries) is positively related to their economic size and inversely related to the trade cost (proxy by geographical distance) between them. However, we depart from previous literature by redefining specific border effects to be measured. By *internal border effect* we denote the number of times a Spanish region trades more with itself than with any another region in the sample. By *external border effect* we denote the number of times a Spanish region trades more with another Spanish region than with a foreign region elsewhere in Europe, controlling for a set of factors.

First, we define our specifications by taking inspiration from some classic papers on the estimation of border effects with sub-national spatial units in Canada and the United States (McCallum, 1995; Anderson and van Wincoop, 2003; Feenstra, 2002, 2004). We will thus for the first time estimate the real flavor of border effects in an EU country as measured with homogeneous spatial units on both sides of the national border (region-to-region, instead of country-to-country or region-to-country). Next, we have our dataset replicate the specifications used to generate previous estimates of region-to-country trade flows in Spain (Gil et al., 2005). We thus highlight for the Spanish case the difference between the benchmark results and the results generated by our new dataset. It is important to note, however, that despite our efforts to keep close to the benchmarks, certain important differences in the datasets will limit the comparability of the results. These differences are laid out in Section 5.

For the sake of brevity, we here define three equations that contain all the models used in this article. They include variables that will be *switched on* or *off* depending on the model in use at a given time. For example, equation [2] formulates a general specification for estimating the external border effect using the inter-regional flows (intra excluded) along with GDPs, distance and other standard control variables:

$$ln\, T_{ijt}^{eu} = \beta_0 + \beta_1 lnY_{it} + \beta_2 lnY_{jt} + \beta_3 External\_Border_{ij}^{eu} + \beta_4 lndist_{ij} + \beta_5 Contig_{ij}^{eu} + \mu_i + \mu_j + \alpha_{i,j} + \gamma_t + \varepsilon_{ij} \quad [2]$$

where lnT_{ij}^{eu} is the logarithm of the flow from region *i* in country *e* to region *j* in country *u* in year *t*. Note that: (a) if *e* = *u* = *Spain* and *i* = *j*, equation [2] will capture intra-regional trade flows for a Spanish region *I*; (b) if *e* = *u* = *Spain* and *i* ≠ *j*, equation [2] will capture inter-regional trade flows for a pair of regions within Spain *ij*; (c) if *e* ≠ *u*, equation [2] will capture inter-regional flows between Spain and another European country in the sample. Since this paper focuses on flows originating in Spanish regions, *e* = *Spain*. The variables lnY_{it} and lnY_{jt} are the logarithms of the nominal gross domestic product (GDP) of the exporting and importing region, respectively. The variable $lndist_{ij}$ is the logarithm of the distance between region *i* and region *j*.

The variable *External_Border* is a dummy that takes the value one for inter-regional flows within Spain (*e* = *u* = *Spain*) and zero otherwise. The anti-log of the parameter associated with this variable measures the size of the *external border effect*.

To capture the positive effect of adjacency, we introduce the dummy variable *Contig*, which takes the value one when trading regions *i* and *j* are contiguous and zero otherwise. This variable conveniently controls for higher inter-regional trade flows between contiguous Spanish regions as well as for the higher concentration of trade between border regions of different countries (Spain-Portugal, Spain-France, Spain-Andorra). It is in line with the results of Lafourcade and Paluzie (2011), who have shown that border regions in countries like France and Spain tend on average to capture larger shares of bilateral trade and FDI flows. The terms μ_i and μ_j correspond to multilateral-resistance fixed effects for the origin and the destination region, respectively. Their inclusion follows Anderson and van Wincoop (2003) and Feenstra (2002, 2004) and is meant to control for competitive effects exerted by the non-observable price index of partner regions and by other competitors. They are also meant to capture other particular characteristics of the regions in question. The variable $\alpha_{i,j}$ is the region-pair effect and γ_t the time fixed effect.

We next define an additional set of models based on equation [3]:

$$ln\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}} = \beta_0 + \beta_1 Internal\_Border_{ij}^{eu} + \beta_2 External\_Border_{ij}^{eu} + \beta_3 dist_{ij} + \beta_4 dist_{ij}^2 + \beta_5 Internal\_Contig_{ij}^{eu} + \beta_6 External\_Contig_{ij}^{eu} + \mu_{i,t} + \mu_{j,t} + \mu^u + \varepsilon_{ijt} \quad [3]$$

where $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$ represents bilateral flows originating in Spanish regions and corrected by the GDPs of the trading regions. We can thus introduce multilateral-resistance terms interacted with time without interfering with monadic variables. Anderson and van Wincoop (2003) have shown that the inclusion of bilateral trade as corrected by unitary income elasticity does not greatly affect the other parameters.

This specification includes the variable *Internal_Border*, which takes the value one when the origin and the destination region are the same (intraregional flows $i = j$) and zero otherwise. It also includes certain refinements in the treatment of distance. It thus includes, apart from the traditional variable $dist_{ij}$, a new variable $dist_{ij}^2$. As in Hillberry and Hummels (2008), Llano-Verduras et al. (2011) and Garmendia et al. (2012), the variable $dist_{ij}^2$ is defined as the square of the distance between trading regions and is expected to capture the non-linear relationship between trade and distance that is observed for kernel regressions in Figure 2. Also in line with these papers, we split the interpretation of these two variables (capturing the negative but non-linear effect of distance on trade) into two parts: (i) a negative and direct effect of distance on trade and (ii) a positive effect for the square of the distance, to capture the high concentration of trade over the shortest distance as observed in the kernel regression. Note also that the *Contig* dummy is also split into two variables: *Internal_Contig* and *External_Contig*. This allows us to consider (simultaneously or independently) the different effects that *adjacency* exerts on trade flows between two contiguous regions in Spain or between a Spanish region and a contiguous foreign one. The terms $\mu_{i,t}$ and $\mu_{j,t}$ correspond to the multilateral-resistance fixed effects for each origin and destination region interacted with time, respectively. It is worth mentioning that, because of their cross-section dataset, the origin and destination fixed effects are included as in Anderson and van Wincoop (2003) and Feenstra (2002). To account for the likely heterogeneity between countries and its effect on the estimate of a single border effect, we have also added a fixed-effect term for each destination country (μ^u).

Finally, as an alternative way to deal with the non-linear relationship between trade and distance, we introduce a flexible approach that controls for changes in the slope of our linear estimation for different "segments" of the sample, these segments corresponding to different distances traveled by trucks. Although purely non-parametric techniques such as kernel regression offer a certain flexibility, they cannot quantify the border effects under discussion. As we will see in the next section, this new approach generates different results from those of log-log linearization or the square of distance. In our view, the variation is due to the differing capacities of the alternative strategies to deal with the *fractal* non-linear relationship shown in Figure 2, which repeats itself at different levels of aggregation, perhaps as flows cross certain thick borders[8]. For each regression using this approach we proceed as follows[9]: (1) we rank the whole sample by increasing distance; (2) we divide the entire range of distance traveled (max-min distance observed in the sample) into “segments” (stretches). For purposes of rigor, we define the “segments” in three alternative ways:

i. **“Naïve”:** The first way simply divides the entire range of actual distance traveled into four stretches of equal length (in kilometers). We call it “naïve” because it ignores the expected higher intensity of flows over the shortest distance.
ii. **“Fibonacci”:** The second way follows the *Fibonacci sequence*, a “magical” mathematical relation that appears in several natural phenomena (the reproduction of rabbits, the internal structure of sunflowers, etc.). The sequence has been used in architecture and in certain fields of economics and finance but, to the best of our knowledge, never before in trade. One benefit of the sequence is that it produces “segments” of increasing length. Another is that the sequence, although completely exogenous, fits perfectly with the non-linear intensity of trade at the nearest distance, dividing the entire range of distance as follows: first stretch: 8% of distance; second stretch: 8%; third stretch: 17%; fourth stretch: 25%; fifth stretch: 42% (100% in total).
iii. **“Quartile”:** The third way assures an equal distribution of the number of observations per segment. It arranges them into quartiles of observation distribution, ranked by distance traveled.

[8] Two examples of thick borders (i.e., administrative borders coinciding with specific forces that cause considerable agglomeration of trade at a short distance) are: (a) Internal borders defining large metropolitan areas; these may coincide with the space where the forces of economic agglomeration around cities are at work, causing a great volume of intra- and inter-regional flow between contiguous regions. (b) International frontiers, coinciding with disproportionate divisions in terms of legal, cultural, historical and political barriers to trade.

[9] Note that segmentation of the sample by range of distance traveled varies for specifications that estimate *internal border effects* (subsample excluding inter-national flows) or focus on *external border effects* (subsample excluding intra-regional flows).

This novel strategy is formally expressed in equation [4]:

$$\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}} = \beta_0 + \beta_1 Internal\_Border_{ij}^{eu} + \beta_2 External\_Border_{ij}^{eu}$$

$$+\mathrm{STRETCH}^s * lndist_{ij} * \theta \; + \beta_4 Internal\_Contig_{ij}^{eu} + \beta_5 External\_Contig_{ij}^{eu} + \mu_{i,t} +$$

$$+\mu_{j,t} + \mu^u + \varepsilon_{ijt} \qquad [4]$$

$\mathrm{STRETCH}^s * lndist_{ij}$ denotes the interaction between the log of the distance and a matrix STRETCH, which contains a set of dummy variables identifying each "segment". By including such interactions, we essentially introduce a set of "semi-dummy" variables, where $lndist_{ij}$ replaces the value one of a normal dummy for the corresponding stretch. θ is a matrix containing the coefficients for each distance stretch. Superscript s indicates the three alternative ways of splitting the sample (Naïve, Fibonacci, Quartile). The rest of the variables are the same as those used in previous specifications.

Having defined all the variables and specifications, we now briefly explain the models used for our empirical analysis and the ways they include our variables and specifications. The estimation methods and data types used for the first set of models are as similar as possible to those of the benchmarks. For models 1 and 2 (M1 and M2), the estimation of equation [2]—as in McCallum (1995) and Anderson and van Winkoop (2003)—is based on cross-section datasets (2007) for region-to-region flows and takes into account only non-zero values (zero values represent 42% of our sample). Estimates for the other models are based on equations [3] and [4] (Tables 3 and 4) and use panel data (2004–2007). For comparability, in M3 and M4 we use the same distance measure as previous authors (Gil et al., 2005). It is a weighted average of geodesic distance between the main cities within each region. For models estimating the *external border effect*, intra-regional trade flows are excluded (and *Internal_Border* therefore drops). For models focusing on the *internal border effect*, international flows are excluded (and *External_Border* drops).

Ordinal Least Square (OLS) estimators are used when the gravity equation is applied to a dataset with no zero values. When zeros are included[10], we instead use instead the Poisson

[10] The zero values considered in our dataset correspond to region dyads that had non-zero values in 2004–2007. Zeros corresponding to regions that did not receive any exports from a Spanish region during that period are not considered in our sample. In this paper, therefore, we are modeling only the intensity of flows between regions, not the drivers behind the existence or non-existence of said flows.

pseudo-maximum likelihood technique (PPML). It was Santos Silva and Tenreyro (2006) who proposed using the PPML approach, which also sorts out Jensen's inequality (note that the endogenous variable is in levels) and produces unbiased estimates of the coefficients by solving the heteroskedasticity problem. For models using the PPML estimator, the independent variable is introduced in levels. For M3 and M4—as in Gil et al. (2005)—we use a panel random effect estimator (REM) with time fixed effects.

5. Results

5.1. Revisiting Previous Estimates

We begin our empirical analysis by revisiting some classic specifications for the estimation of border effects, so as to test the performance of our new dataset against them. This will afford us some measure of comparability with previous results and thus allow us to determine which of our results derive from new specifications and which from our dataset itself. Table 2 lays out the results for four models.

<< Table 2 about here >>

The first two models (M1 and M2) were inspired by McCallum (1995), Anderson and van Wincoop (2003) and Feenstra (2002, 2004)[11]. As reported in the first column (M1), when fed our dataset the McCallum-like specification generates an external border effect of 15 (vs. McCallum's 22 for Canada-US!). The coefficients and signs for the rest of the variables align with expectations. There is, however, a slightly lower coefficient for GDPs than the normal values in our benchmarks, which use all trade flows and not just truck deliveries. Similarly, the figures for model 2 (M2) were generated by our novel dataset and a specification similar to that defined by Anderson and van Wincoop (2003) and Feenstra

[11] Before comparing results, we should point out some relevant differences between our dataset and the dataset used by these authors: (i) We must emphasize that figures obtained for external border effects in a country like Spain in its trade with eight European countries can hardly be compared with the figures for Canada and the US. (ii) Our distance variable for intra- and inter-national flows measures the actual distance travelled by trucks delivering commodities, whereas the distance used by McCallum (and in papers published thereafter) was either the linear distance between the main cities in each province and state or the weighted distance. (iii) In our sample we consider eight different "international borders", two of them between Spain and three contiguous countries (Andorra, France and Portugal), whereas McCallum and all the subsequent articles replicating his work with similar datasets considered only one "international border": between Canada and the United States.

(2002, 2004). The results once again align with expectations. As in the benchmark papers, we find a significant decrease in the external border effect (now of factor 4) when multilateral resistance terms are taken into consideration.

M3 and M4 report the results generated by a specification similar to the one used by Gil et al. (2005) but with our novel dataset. To reduce differences, we aggregate our region-to-region dataset to the data structure used in their paper—flows between each Spanish region to the rest of Spain (ROS)—and use their distance measure[12]. Like them, we also omit data on intra-regional flows and focus on the estimation of the *external border effect*. Table 2 reports the results for model 3 (M3) and model 4 (M4). M3 suggests negative elasticity for a distance of –1.844 and an external border effect of 38 [exp(3.638)] for Spanish exports. Similarly, Gil et al. (2005) obtained negative elasticity at a distance of –1.28 when using GDPs and at a distance of –1.26 when using Population and Surface (columns (1) and (3) in Table 1 of their paper). With these two specifications, which do not control for contiguity, they obtained an *external border effect* of 20 [exp(2.99)] for exports and of 24 [exp(3.18)] for imports. With model 4 (M4), where they controlled for contiguity (as we have done), they obtained a lower negative coefficient for distance (–0.88, vs. –1.299 in our estimates). However, they obtained a lower external border effect for exports than we have[13].

5.2. Focus on the Non-Linear Relationship between Trade and Distance

In this section we analyze the main results for the sixteen augmented models that use equations [3] and [4] and the region-to-region dataset. For comparability with previous papers, external and internal border effects are estimated separately.

<< Table 3 about here >>

[12] Some important differences nevertheless hold. For example: (1) Gil et al. (2005) used a different database on inter-regional trade flows within Spain for 1995–1998, used international flows by all transport modes for twenty-seven OECD countries, and included Spain's two island regions. Our estimate uses data for 2004–2007, considers only inter-regional and inter-national flows by Spanish trucking to eight European countries, and excludes the islands. (2) Gil et al. (2005) used trade flows and GDPs in real terms; we use them in current terms.

[13] In our view, our larger border effect for exports can be explained by the differences between the two datasets. Gil et al. (2005) used total flows (not just truck deliveries) and a wider range of countries. Although the number of deliveries by Spanish trucks could be taken as representative of all internal trade flows (trucks accounting for more than 90% of Spain's internal transport flows), the international truck deliveries in our sample fall far short of the total trade considered in their all-modes sample for twenty-seven OECD countries. (Again, we consider only Spanish trucking flows to eight European countries.)

Table 3 reports results for a first set of models estimating the *external border effect*. Similarly, Table 4 reports the corresponding results for the *internal border effect*. All of these models use the *corrected trade* flows $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$ as an endogenous variable as well as all the fixed effects described above. However, each uses a different treatment of the distance variable. M5 and M11 include the endogenous variable and the distance in logs (OLS without zero flows). M6 and M12 include zero flows and use the PPML estimator. Thus the endogenous variable is expressed in levels and distance in logs. In M7 and M13, *distance* and the *square of distance* are included in levels. Finally, to shed more light on the non-linear relationship between trade and distance, Table 3 (M8–M10) and Table 4 (M14–M16) report the corresponding results for six alternative models based on our alternative strategy (equation [4]), which segments the sample three ways by trucking distance. This procedure estimates the elasticity of distance in each interval. Note that in these models the distance variables for each "stretch" are also expressed in logs.

First, we analyze the results obtained for the *external border* (Table 3). The first three models generate significant coefficients with the expected signs for all variables except *External_Contig* in M5 and M6. This result suggests that the difference in the intensity of trade between a Spanish region and a foreign border region, on the one hand, and between non-adjacent Spanish regions, on the other, is non-significant, whether the intensity is higher or lower. Note, in fact, that the coefficient for the *Internal_Contig* variable is positive and significant. Moreover, the results for distance variables that control for the non-linear relationship between trade and distance in M7 suggest that distance acts as a clear impediment to trade (negative coefficient for *Dij*), but an impediment that tapers off as distance increases (positive coefficient for the square of distance). As for the *external border effect*, these three models reach a similar factor of 6, which stands up robustly to alternative specifications, subsamples, estimation procedures and treatments for the non-linear relationship between trade and distance. This persistent *external border effect* (6) is slightly larger than that obtained by Llano-Verduras et al. (2011) with region-to-country [3.3 = exp(1.2)] and province-to-country [4.9 = exp(1.6)] data. Note that the papers use different datasets but similar specifications for distance and the same estimation procedures as in M6 and M7 (PPML)[14]. The *external border effect* for Spanish exports (6) is considerably smaller than the 38 and 63 obtained when the same dataset is applied in a region-to-country aggregated format (M3 and M4), as in Gil et al. (2005).

Table 3 reports promising results for models M8 to M10, which employ our new controls for non-linearity. The coefficients for *STRETCH**ln$dist_{ij}$ in each of the segments are negative and highly significant for these three alternative models. More interestingly, in M8 (Naïve) and M10 (Quartile) the negative elasticity for each stretch decreases, which is consistent with a segmentation where the distance variable is shorted in increasing order. In M9 (Fibonacci), however, the negative elasticity of distance increases in the first two segments (from –0.966 to –1.020) and decreases thereafter (from –1.020 to –0.907). As for the *external border effect*, the three alternative procedures for segmenting the sample reach very similar positive and significant coefficients, in line with those obtained in the previous models. Finally, it is interesting to note that in M9 and M10—i.e., the models where every stretch is of a different length—the coefficient for *Internal_Contig* becomes non-significant. This suggests that when the non-linear relationship is controlled for by segmentation of the sample into stretches of increasing length (and intensity of trade), the alternative control for the higher intensity of trade over the shortest distance (*Internal_Contig*) becomes redundant.

<< Table 4 about here >>

We now focus on our results for the *internal border effect*, reported in Table 4. As previously stated, the structure of the table and the models is equivalent to that in Table 3. The most surprising result is the non-significance of the *internal border effect* in all but one model, M13, where it reaches a factor of 10. This border effect is greater than the one obtained by Garmendia et al. (2012) with province-to-province data, OLS [3.7 = exp(1.31)] and PPML [2.4 = exp(0.88)] procedures and similar specifications for the distance variable (square of distance). Now, as in Llano-Verduras et al. (2011), the contiguity dummy

[14] In Llano-Verduras et al. (2011), the external border effect was computed with PPML procedures. Simply by disaggregating the spatial unit of the exporting area in Spain, they were able to obtain a significant reduction in the border effect. That said, the squared-distance term was non-significantly different from zero with region-to-country data, but only at the province-to-country level. When our novel dataset is used with full disaggregation (zeros included and region-to-region), the external border effect slightly decreases (1.869 in M6 versus 1.852 in M7) and the square of the distance has a positive and significant coefficient (0.767 in M7). We believe these results in part from the splitting of importing countries into their corresponding regions as well as for the difference in the datasets used in the two papers.

becomes non-significant and negative for all models. It is worth mentioning that the *internal border effect* increases when zero flows are included and the PPML is used but almost disappears when segmented distance is used. Interestingly, unlike the corresponding models in Table 3, the last three models in Table 4 (M14–M16) show that negative elasticity for each stretch always decreases. When it comes to levels, moreover, the elasticity for each segment in the "national flows" in Table 4 is always higher than its Table 3 counterpart, "interregional national-international flows". In fact, the highest negative elasticity in Table 3 (stretch 1, Model 10, Quartile: –1.084) is lower than the lowest in Table 4 (stretch 4, Model 14, Naïve: –1.100). This suggests a sharper decrease in trade as distance increases for flows within the country (because of both intra- and inter-regional flows between contiguous regions of the same country). Note also that the last two models (M15 and M16), which consider stretches of increasing length, best capture the non-linear relationship (higher elasticity of distance in the first segments) and generate the lowest *internal border effect* (1). However, it is also remarkable that the *external border effect* almost remains fixed at a factor of 7, which is equivalent to the effects obtained with the other specifications.

<< Table 5 about here >>

To bring home the previous results, Table 5 summarizes the main features of the three alternative segmentations and provides measures of overall fitness for each segment. For the sake of clarity, results for *external* and *internal border effect* estimates are reported separately. Several points are worth mentioning: (i) The three sequences have been defined not by volume or nature of trade but, instead, by distance range and number of observations (zero flows included). Thus the *Fibonacci* and *Quartile* sequences consider segments of different length. (ii) The percentage of zero values for each stretch and each criterion is different. Zero values are highly concentrated in the longest trips (mainly international flows). (iii) For a complementary view, we show fitness obtained with the models run separately for each segment of the sample[15]. Table 5 reports the R^2 for regressions that use (lnD_{ij}) or (D_{ij}, D_{ij}^2): i.e., for the counterparts to the specifications used in M6 and M7 (*external border*) and in M12 and M13 (*internal border*). Note that, although the three

[15] Note that the results in Tables 3 and 4 consider the whole sample and use the strategy for equation 4, where each segment is controlled by a semi-dummy obtained through the interaction of a dummy and the distance variable.

alternative segmentation criteria generate the same R^2 (lnD_{ij}) and $R^2(D_{ij}, D_{ij}^2)$ for the whole sample (TOTAL column), the quality of the fit is different for each segment and sequence. Throughout the sample (TOTAL), R^2 is always higher with lnD_{ij} (89%) than with D_{ij} or D_{ij}^2 in levels (84%). In the first part of the table (External Border), the *Fibonacci* sequence (followed by *Quartile*) shows the best fits when the model is regressed for the last subsamples (largest distances) and generates the highest R^2 (lnD_{ij}) and $R^2(D_{ij}, D_{ij}^2)$. Conversely, although the *Naïve* sequence performs well for the first two stretches, it fails for the last two. In the second panel of the table (Internal Border), the *Fibonacci* sequence just overcomes the others in the third stretch, whereas *Quartile* shows the best fits for the rest of the subsamples.

At this point, it is worthwhile to sum up our results, which might have something to do with the nature of the two border effects considered here. On the one hand, the *internal border effect*, far from being explained by external barriers to trade (division or fragmentation), seems most closely related to the economics of agglomeration around metropolitan areas, as well as to the spatial spillover of the strongest regions and their neighbors. It thus seems sensitive mostly to mismeasurement, spatial-unit use (optimal unit being MAUP) and aggregation bias. The *external border effect*, on the other hand, seems to harder to budge (Wei, 1996). First, region-to-region international flows lead to lower external borders (15 and 4 for M1 and M2, respectively) than do region-to-country datasets (38 and 63 for M3 and M4). However, even when we include zero flows (which tend to increase the external border, since most zero flows correspond to international flows) and control for the non-linear relationship of trade, we obtain a positive and significant factor of 6 or 7. Finally, according to our results, it is not clear that log-log linearization or the square of the distance improves the treatment of the observed non-linearity, whereas our strategy of segmenting the sample (especially the two procedures that produce stretches of increasing length) does seem to improve its treatment. Nevertheless, our results show larger variations in the elasticity of distance (by segment) and in the role played by (external and internal) contiguity than the *border effects* themselves.

5.3. Results by Importing Country

Before we conclude, we would like to discuss in greater detail the nature of this *persistent external border effect*, taking each country separately. As stated above, this will shed new light on the differences between the national borders of contiguous and non-contiguous countries. It will also allow us to infer the heterogeneous level of trade integration (at least on the export side) between Spanish regions and the regions of Spain's main European partners.

<< Table 6 about here >>

Table 6 reports our results for the *external border effect* of each importing country with M17, M18 and M19. Note that intra-regional flows and the Internal-Border are not included in the analysis. The *external border effect* is now expressed negatively, indicating how many fewer times a Spanish region exports to a non-adjacent region in France than to a non-adjacent Spanish region, ceteris paribus. The results are ranked by increasing order of *external border effect* as obtained with M18 and M19. The lowest *external border effects* are obtained for Andorra (7), Belgium (9), Portugal (10), France (10) and Germany (10), followed by the Netherlands, the UK and Italy (all with 12).

If we consider this *external border effect* as a measure of integration between Spanish regions and the regions of these eight European partners—with size, bilateral distance and contiguity previously controlled for—it is remarkable to find the highest levels of integration not only with the regions of the nearest countries (Andorra, Portugal and France) but also with the regions in Belgium and Germany. This complements results in papers that used region-to-country flows but centered on the performance of key Spanish regions like Cataluña and the País Vasco[16]. Belgium's border effect, lower than that for countries of similar size and accessibility by road (i.e., the Netherlands), may be connected with the country's specific features, such as the location of seat of the EU government in Brussels.

Finally, some brief comments on our other variables. Segmented distance performs as expected, with a decreasing negative elasticity for the longest stretches. Only in M18-Fibonacci do the first two segments show a lower negative elasticity than the next. Moreover, to control for the enhancing effect of border regions (Lafourcade and Paluzie,

[16] Note also that the results are not fully comparable, because of notable differences in the data type and specifications used in each paper. Thus Ghemawat et al. (2010), using total trade flows (exports + imports by all transport modes), found a low and shrinking external border effect for Catalonia in its trade with France. Another case in point is Gil-Pareja et al. (2006), who also found—using regional balance-of-payment data and region-to-country data—that the external border effect for Basque Country exports was lower for trade with countries such as Portugal, Belgium, Germany, Finland and the Czech Republic than it was for trade with France. They also found greater border effects for Basque Country exports to Italy (38) and the UK (41), even when all trade flows (and not just trucking) were considered.

2011), we split the *Contiguity* variable into a set of alternative dummies: for Spanish regions contiguous with a foreign region (*Internal_Cont*) and for contiguous foreign regions in any of Spain's three neighbors (*External_Cont_FR*, *External_Cont_PT*, *External_Cont_AD*). We should mention that, when the non-linear relationship between trade and distance is controlled for by segmented distance, all contiguity variables become non-significant. As stated above, these results suggest that the border effect and the contiguity dummy usually absorb part of the non-linearity. They also indicate that the clear concentration of international flows in the shortest distance (border regions beyond the national border) shown by Figure 2 is not statistically different from the concentration observed in a typical inter-regional flow within the country. In conclusion, although bordering regions of contiguous countries are more integrated than are other regions foreign to each another, they are still less integrated than any pair of regions within the same country.

6. Conclusions

In this article we aim to measure the internal and external trade integration of Spanish regions by quantifying the *external* and *internal border effects* in Spain, taking into account intra- and inter-national trade between Spanish regions (NUTS 2) and the regions of Spain's eight main European partners. Lack of information on inter-regional flows, both domestic and international, has hitherto impeded pristine estimates of *border effects* like those obtained for Canada and the US by McCallum (1995), Anderson and van Wincoop (2003) and others. Until now, *external border effects* in Europe have been computed with country-to-country or region-to-country flows. We can thus suppose there to have been a loss of relevant information, because of aggregation and the use of non-homogeneous spatial units. The computation of distance as well, both for intra-national trade and for region-to-country dyads, has seemed a source of bias in previous border-effect estimates.

In this paper we have made use of a novel dataset for inter-regional trade flows by Spanish trucking, including intra-national and inter-national flows and considering actual distance for the shipments. From this starting point we have borrowed classic specifications previously used to compute *external* and *internal border effects* both between Canada and the US and between Spain and other countries. As our results attest, the new dataset generates similar but not identical figures when channeled into these classic specifications. For one thing, we find a higher level of integration (i.e., a lower external border effect) between our region pairs than previous authors have found between Canadian provinces and US states (15 vs. 22). When we aggregate our dataset to a format of region-to-country Spanish flows (Gil et al., 2005), we obtain a greater border effect than the benchmark paper found. However, when we use the dataset with full disaggregation (with zeros and region-to-region) and some new estimation procedures, external and internal border effects decrease or become non-significant. We have also developed a new strategy to deal with the *fractal non-linear relationship between trade and distance*. Namely, we segment the sample, considering alternative stretches of the distance variable. With this approach the internal border effect vanishes, while the external border remains a positive, significant factor of 7. Although this positive external border effect persists in all of our models, we obtain an important reduction simply by using homogeneous spatial units on both sides of a given national border, even with Nuts 2. Finally, we have repeated the analysis considering country-specific border effects and found that the level of integration between Spanish regions and their main partners (lower border effects) depends mainly on proximity (except in the case of Belgium).

7. References

Anderson, J.E. and Van Wincoop, E. (2003). “Gravity with Gravitas: A Solution to the Border Puzzle”, *American Economic Review*, **93**, 1, 170-192.

Anderson, M., and S. Simth (1999). Do National Borders Really matter? Canada-US Regional Trade Reconsidered. *Review of International Economics* 7 (2): 219-227.

Brakman S., Garretsen H. van Marrewijk C. (2009): The New Introduction to Geographical Economics. Cambridge University Press.

Chen, N. (2004). “Intra-National Versus International Trade in the European Union: Why Do National Borders Matter?”, *Journal of International Economics*, **63**, 1, 93-118.

Combes, P. P., Lafourcade, M. and Mayer, T. (2005) "The trade-creating effects of business and social networks: evidence from France", *Journal of International Economics*, **66**: 1–29.

Daumal M, Zignago S (2008) Border effects of Brazilian states. Working Papers 2008-11, CEPII Research Center

Deardorff AV (1998). “Determinants of bilateral trade: does gravity work in a neoclassic world?” In J. Frankel JA (ed) *The Regionalization of the World Eeconomy*. University of Chicago Press, Chicago

Djankov S, Freund C (2000). “Disintegration and trade flows: evidence from the Former Soviet Union”, Policy Research Working Paper Series 2378. *The World Bank*

Evans, C.L. (2003). "The Economic Significance of National Border Effects", *American Economic Review*, **93**, 4, 1291-1312.

Feenstra R (2002). “Border effect and the gravity equation: consistent methods for estimation”, *Scottish Journal of Political Economy*, 49(5), 1021–1035

Feenstra R., (2004) “Advanced International Trade” *Princeton University Press*.

Gallego N. and Llano C. (2012). “Towards a region-to-region international trade database for the EU: methodology of estimation and first results for the Spanish case”, Working paper presented in the NARSC 2011. Miami.

Garmendia, A., Llano-Verduras, C. and Requena-Silventre, F. (2012) “Network and the disappearance of the intranational home bias,” *Economic Letters* 116, 178-182.

Ghemawat, P., Llano-Verduras. C., Requena-Silventre, F., (2010). “Competitiveness and interregional as well as and international trade: The case of Catalonia”, *International Journal of Industrial Organization,* **28**, 415–422

Gil-Pareja, S., Llorca-Vivero, R., Martínez Serrano J.A. and Oliver-Alonso, J. (2005). "The Border Effect in Spain", *The World Economy*, **28**, 11, 1617-1631.
Gil-Pareja, S., Llorca-Vivero, R., Martínez Serrano J.A. (2006): The Border Effect in Spain: The Basque Country Case. *Regional Studies*, **40.4**, 335–345.
Head, K. and Mayer, T. (2000). "Non-Europe: the causes and magnitudes of market fragmentation in the EU", *WeltwirtschaftlichesArchiv*, **136**, 2, 285-314.
Head, K. and Mayer, T. (2002). "Illusory Border Effects: Distance mismeasurement inflates estimates of home bias in trade", *CEPII Working Paper 2002-01*.
Heinemeyer, H., Chulze, M.-H. and Wolf, N. (2008), "Endogenous Borders? The Effects of New Borders on Trade in Central Europe 1885-1933", CESifo Working Paper No. 2246
Helble (2007). "Border Effect Estimates for France and Germany", *Review of World Economics*, Vol. 143 (3)
Helliwell, JF (1996). "Do national borders matter for Quebec's trade?", *Canadian Journal of Economics*, 29(3), 507–522.
Helliwell, JF (1997). "National Borders, Trade and Migration", *Pacific Economic Review* 3 (3): 507-522
Helliwell JF (1998). "How much do national borders matter?" *Brookings Institution Press, Washington, DC.*
Helliwell, J.F. and Verdier, G. (2001). "Measuring internal trade distances: a new method applied to estimate provincial border effects in Canada", *Canadian Journal of Economics*, **34**, 4, 1024-1041.
Helpman, E., Melitz, M and Rubinstein, Y. (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes", *Quarterly Journal of Economics*, **123**, 2, 441-487.
Hillberry, RH (1998). "Regional Trade and "the Medicine Line": The External border effect in US Commodity Flow Data", *Journal of Borderland Studies* 13: 1-17.
Hillberry, RH (2002). "Aggregation bias compositional change, and the border effect", *Canadian Journal of Economics,* **35**(3), 517–530.
Hillberry RH, Hummels D (2003). "Intranational home bias: some explanations", *Review of Economics and Statistics*, **85**(4), 1089–1092.
Hillberry, R. and Hummels, D. (2008). "Trade responses to geographic frictions: A decomposition using micro-data", *European Economic Review*, **52**, 3, 527-550.
Lafourcade, M. and Paluzie, E. (2011) "European Integration, Foreign Direct Investment (FDI), and the Geography of French Trade", *Regional Studies*, **45**: 4, 419 — 439.
Llano-Verduras C.; Minondo A., Requena-Silvente F. (2011). "Is the Border Effect an Artefact of Geographical Aggregation?" *The World Economy*, **34**, 10, 1771–1787.
Llano, C., Esteban, A., Pulido, A., Pérez, J (2010): "Opening the Interregional Trade Black Box: The C-intereg Database for the Spanish Economy (1995-2005)". *International Regional Science Review*, **33**, 302-337.
McCallum, J. (1995). "National Borders Matter: Canada-US. Regional Trade Patterns", *American Economic Review*, **85**, 3, 615-623.
Millimet D, and Osang Th (2007). "Do state borders matter for US. intranational trade? The role of history and internal migration", *Canadian Journal of Economics*, **40**(1), 93–126.
Minondo A. (2007). "The disappearance of the border barrier in some European Union countries' bilateral trade", *Applied Economics*, **39**, 119-124.
Nitsch, V. (2000). "National borders and international trade: evidence from the European Union", *Canadian Journal of Economics*, **33**, 4, 1091-1105.
Nitsch, V., (2002). "Border effects and border regions: lessons from the German unification". Banakgesellschaft, Berlin.
Obstfeld, M. and Rogoff, K. (2000). "The Six Major Puzzles in International Macroeconomics. Is there a Common Cause?", *NBER Working Paper 7777*, National Bureau of Economic Research, Cambridge, MA.
Okubo, T., (2004) "The border effect in the Japanese market: a gravity model analysis". *Japan International Economics*, **18**:1–11
Requena, F. and Llano, C. (2010) "The border effects in Spain: an industry level analysis", *Empirica*, 37, 455-476.
Shultze, M.S. and N. Wolf (2008), "On the origins of border effects: insights from the Habsburg Empire", *Journal of Economic Geography*, 1-20.
Silva, J. and Tenreyro, S. (2006). "The log of gravity", *Review of Economics and Statistics*, **88**, 4, 641-658.
Wei, S. (1996). "Intra-National Versus International Trade: How Stubborn are Nations in Global Integration?", *NBER Working Paper 5531*.
Wolf, H.C. (1997). "Patterns of Intra- and Inter-State Trade. NBER Working Paper 5939. National Bureau of Economic Research, Cambridge, Mass.
Wolf, H.C. (2000). "Intranational home bias in trade", *Review of Economics and Statistics*, **82**, 4, 555-563.

TABLES

Table 1. Selected Papers on External Border Effect for North America, OCDE, Europe and Spain, Classified by Data Type and Spatial Unit.

Paper	Country	Sectoral analysis	Time period	External border effect
Region-to-region				
1995. McCallum	Canada-United States	No	1988	22
1996. Helliwell	Canada-United States	No	1988–1990	22
1998. Hillbery	Canada-United States	No	1993	20
2001. Helliwell	Canada-United States	No	1991–1996	15–10
2002. Head & Mayer	United States (Wolf, 1997, 2000)	Yes	1997	11
Country-to-country				
1996. Wei	OCDE	No	1982–1994	10-2.6
1997. Helliwell	OCDE	No	1996	13
2000. $Nitsch_a$	EU-10	No	1979–1990 1983–1990	7–10
2000. Head & Mayer	EU-9 EU-12	Yes Yes	1976–1995 1993–1995	30-11 13
2004. Chen	EU-7	Yes	1996	6
Region-to-country				
1999. Anderson & Smith	Canada-United States	No		12
2005. Gil et al.	Spain (17 regions), Rest of Spain(*) and OECD-27	No	1995–1998	21
2003. Minondo	Basque Country, Rest of Spain(*), 201 countries	No	1993–1999	20–26
2007. Helble	France, EU-14 Germany, EU-14	No	2002	8 3
2010. Requena &Llano	Spain (17 regions) OECD-28	No Yes	1995 & 2000	13
2010. Ghemawat et al.	Catalonia, Rest of Spain(*), OECD	Yes	1995–2006	55
2011. Llano-Verduras et al.	Spain (17 regions; 50 provinces, OECD)	No	2000 & 2005	40

(*) Rest of Spain considered as a country, with total exports computed from one Spanish region to the rest of Spain (ROS). The purpose of this aggregation is to measure external border effects when region-to-region data is not available.

Table 2. Estimations with Classic Specifications from Previous Papers. Based on Eq. [2].

Model:	M1	M2	M3	M4
Estimation method	OLS	OLS	REM	REM
Reference paper:	Mc-1	AvW-1	Gil et al.-4	Gil et al.-5
VARIABLES	$Ln(T_{ijt})$	$Ln(T_{ijt})$	$Ln(T_{iJt})$	$Ln(T_{iJt})$
Ln(GDPit)	0.566***	0.821***	0.998***	0.958***
	(0.0389)	(0.129)	(0.0921)	(0.0891)
Ln(GDPjt)	0.554***	0.624***	0.668***	0.619***
	(0.0439)	(0.157)	(0.124)	(0.122)
Ln(Dij)	–1.181***	–1.263***	–1.844***	–1.299***
	(0.0495)	(0.0447)	(0.284)	(0.315)
External Border	2.687***	1.290***	3.638***	4.143***
	(0.0937)	(0.354)	(0.261)	(0.291)
Contig				1.167***
				(0.289)
Constant	–1.952	–8.266*	12.26***	9.432***
	(1.505)	(4.449)	(1.843)	(1.990)
External border = $\exp(\beta_3)$	15	4	38	63
Observations	896	896	436	436
Period	2007	2007	2004–2007	2004–2007
R^2	0.774	0.839	-	-
Pseudo R^2	-	-	0.7558	0.7679
Multilateral resistance	NO	YES	NO	NO
Time fixed effect	NO	NO	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

**Table 3. Alternative Estimates for *External Border Effects*.
M5-M7 Based on Eq [3], M8–M10 on Eq [4].**

	M5	M6	M7	M8-Naïve	M9-Fibonacci	M10-Quartile
	OLS	PPML	PPML	PPML	PPML	PPML
VARIABLES	Ln(Tijt corr)	Tijt corr	Tijt corr	Tijt corr	Tijt corr	Tijt corr
$Ln(D_{ij})$	–1.032***	–0.927***				
	(0.0728)	(0.103)				
D_{ij}			–2.775***			
			(0.353)			
D_{ij}^2			0.767***			
			(0.104)			
$Ln(D_{ij}$ stretch1)				–1.028***	–0.966***	–1.084***
				(0.154)	(0.200)	(0.177)
$Ln(D_{ij}$ stretch2)				–1.025***	–1.018***	–1.030***
				(0.142)	(0.181)	(0.165)
$Ln(D_{ij}$ stretch3)				–0.975***	–1.020***	–0.994***
				(0.137)	(0.167)	(0.161)
$Ln(D_{ij}$ stretch4)				–0.931***	–0.960***	–0.950***
				(0.133)	(0.156)	(0.157)
$Ln(D_{ij}$ stretch5)					–0.907***	
					(0.150)	
External_Border	1.814***	1.869***	1.852***	1.909***	1.995***	1.980***
	(0.410)	(0.281)	(0.299)	(0.282)	(0.284)	(0.299)
Internal_Contig	0.308***	0.550***	0.901***	0.416*	0.213	0.341
	(0.0988)	(0.203)	(0.250)	(0.236)	(0.285)	(0.307)
External_Contig	0.491***	0.163	0.343	0.0334	-0.0926	0.00165
	(0.159)	(0.325)	(0.328)	(0.338)	(0.347)	(0.365)
Constant	–23.63***	–24.90***	–29.40***	–24.31***	–24.50***	–24.03***
	(0.640)	(0.717)	(0.336)	(1.000)	(1.123)	(1.154)
External Border = $\exp(\beta_2)$	6	6	6	7	7	7
Observations	3,631	6,333	6,333	6,333	6,333	6,305
R^2	0.810	0.888	0.842	0.890	0.891	0.889

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All regressions include an "origin by year", "destination by year" and "destination country" fixed effect.
Tijt_corr = $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$

Table 4. Alternative Estimates for the *Internal Border Effect*. M11–M13 Based on Eq [3], M14–M16 on Eq [4].

	M-11	M-12	M-13	M14-Naïve	M15-Fibonacci	M16-Quartile
	OLS	PPML	PPML	PPML	PPML	PPML
VARIABLES	Ln(Tijt corr)	Tijt corr	Tijt corr	Tijt corr	Tijt corr	Tijt corr
$Ln(D_{ij})$	–1.274***	–0.926***				
	(0.0596)	(0.209)				
D_{ij}			–5.397**			
			(2.459)			
$D_{ij}2$			3.625*			
			(1.861)			
Ln(D_{ij} stretch1)				–1.266***	–1.568***	–1.261**
				(0.307)	(0.532)	(0.502)
Ln(D_{ij} stretch2)				–1.189***	–1.482***	–1.224***
				(0.283)	(0.467)	(0.450)
Ln(D_{ij} stretch3)				–1.151***	–1.449***	–1.172***
				(0.267)	(0.431)	(0.424)
Ln(D_{ij} stretch4)				–1.100***	–1.356***	–1.133***
				(0.256)	(0.394)	(0.403)
Ln(D_{ij} stretch5)					–1.303***	
					(0.371)	
Internal_Border	0.184	0.920	2.291***	0.469	0.243	0.377
	(0.214)	(0.732)	(0.784)	(0.853)	(0.927)	(1.200)
Internal_Contig	–0.0114	–0.284	–0.0527	–0.0856	–0.0850	–0.206
	(0.0820)	(0.307)	(0.468)	(0.393)	(0.362)	(0.460)
Constant	–20.24***	–22.43***	–26.55***	–20.94***	–19.78***	–20.86***
	(0.392)	(1.341)	(0.796)	(1.732)	(2.472)	(2.695)
Internal Border = exp(β_1)	1	3	10	2	1	1
Observations	899	900	900	900	900	876
R^2	0.887	0.919	0.877	0.931	0.930	0.929

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All the regressions include an "origin by year", "destination by year" and "destination country" fixed effect.
Tijt_corr = $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$

Table 5. Characterization for three Alternative Segmentations.

External Border							
Naive	Stretch Km. (thousands)	0.0512–0.8134	0.8135–1.5756	1.5757–2.3378	2.3379–3.1		TOTAL
	Range	25%	25%	25%	25%		100%
	obs. (%)	16%	18%	18%	4%		56%
	zeros (%)	1%	13%	24%	5%		44%
	R^2 (lnDij)	**87%**	**79%**	**23%**	**4%**		89%
	R^2 (Dij, Dij^2)	**86%**	**79%**	**23%**	**6%**		84%
Fibonacci	Stretch Km. (thousands)	0.0512–0.3052	0.3053–0.5593	0.5594–1.0674	1.0675–1.8296	1.8297–3.1	TOTAL
	Range	8%	8%	17%	25%	42%	100%
	obs. (%)	5%	5%	13%	19%	14%	56%
	zeros (%)	0%	0%	2%	23%	18%	44%
	R^2 (lnDij)	**85%**	**77%**	**74%**	**35%**	**13%**	89%
	R^2 (Dij, Dij^2)	**84%**	**77%**	**75%**	**36%**	**14%**	84%
Quartile	Stretch Km. (thousands)	0.0512–1.0624	1.0625–1.590	1.591–1.9651	1.9652–3.1		TOTAL
	Range	33%	17%	12%	37%		100%
	obs. (%)	22%	13%	10%	11%		56%
	zeros (%)	3%	12%	15%	14%		44%
	R^2 (lnDij)	**88%**	**59%**	**38%**	**19%**		89%
	R^2 (Dij, Dij^2)	**86%**	**61%**	**39%**	**19%**		84%
Internal Border							
Naive	Stretch Km. (thousands)	0.0126–0.2896	0.2897–0.5665	0.5666–0.8434	0.8435–1.124		TOTAL
	Range	25%	25%	25%	25%		100%
	obs. (%)	33%	24%	30%	12%		99.9%
	zeros(%)	0%	0%	0.1%	0%		0.1%
	R^2 (lnDij)	**92%**	**61%**	**36%**	**61%**		92%
	R^2(Dij, Dij^2)	**89%**	**60%**	**36%**	**61%**		88%
Fibonacci	Stretch Km. (thousand)	0.0126–0.1049	0.1050–0.1973	0.1974–0 .3819	0.3820–0.6588	0.6589–1.1204	TOTAL
	Range	8%	8%	17%	25%	41%	100%
	obs. (%)	12%	10%	18%	31%	29%	99.8%
	zeros (%)	0%	0%	0%	0.1%	0%	0.1%
	R^2 (lnDij)	**90%**	**72%**	**79%**	**61%**	**51%**	92%
	R^2 (Dij, Dij^2)	**91%**	**72%**	**80%**	**61%**	**51%**	88%
Quartile	Stretch Km. (thousand)	0.0126–0.2236	0.2237–0.4916	0.4917–0.6739	0.6740–1.1204		TOTAL
	Range	19%	24%	16%	40%		100%
	obs. (%)	25%	25%	25%	25%		99.9%
	zeros (%)	0%	0%	0.1%	0%		0.1%
	R^2 (lnDij)	**92%**	**75%**	**59%**	**62%**		92%
	R^2 (Dij, Dij^2)	**90%**	**76%**	**59%**	**63%**		88%

All the regressions include an "origin by year", "destination by year" and "destination country" fixed effect. Tijt_corr = $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$

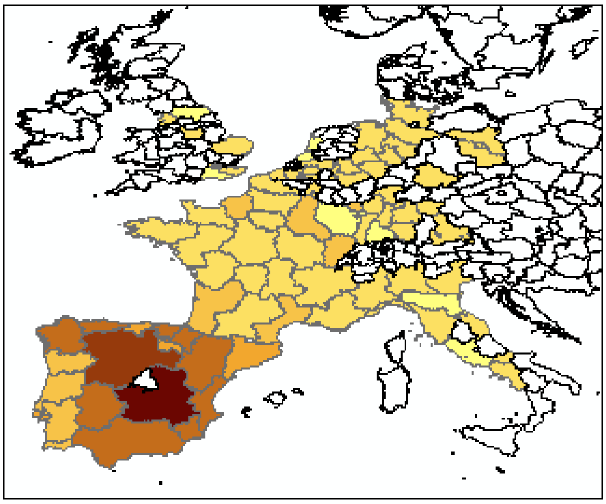
Table 6. External Border Effects by Country. PPML Procedure. Region-to-Region Spanish exports 2004–2007.

VARIABLES	(M17)-Naïve PPML Tijt_corr		(M18)-Fibonacci PPML Tijt_corr		(M19)-Quartile PPML Tijt_corr	
Ln(Dij stretch1)	–1.029***	(0.193)	–0.967***	(0.276)	–1.086***	(0.178)
Ln(Dij stretch2)	–1.025***	(0.177)	–1.019***	(0.248)	–1.031***	(0.166)
Ln(Dij stretch3)	–0.976***	(0.171)	–1.021***	(0.227)	–0.995***	(0.162)
Ln(Dij stretch4)	–0.931***	(0.167)	–0.961***	(0.214)	–0.951***	(0.158)
Ln(Dij stretch5)			–0.907***	(0.206)		
Border_AD	–1.906***	(0.306)	–1.990***	(0.308)	–1.966***	(0.306)
Border_BE	–1.954***	(0.212)	–2.232***	(0.216)	–2.253***	(0.215)
Border_PT	–2.305***	(0.178)	–2.292***	(0.179)	–2.308***	(0.177)
Border_FR	–2.207***	(0.158)	–2.341***	(0.157)	–2.350***	(0.156)
Border_DE	–2.035***	(0.203)	–2.309***	(0.206)	–2.352***	(0.207)
Border_NL	–2.179***	(0.219)	–2.463***	(0.223)	–2.500***	(0.222)
Border_UK	–2.226***	(0.215)	–2.469***	(0.218)	–2.525***	(0.217)
Border_IT	–2.246***	(0.210)	–2.488***	(0.214)	–2.551***	(0.214)
Internal_Cont	0.416	(0.307)	0.213	(0.360)	0.340	(0.309)
External_Cont_FR	0.0617	(0.507)	–0.0429	(0.519)	0.0602	(0.513)
External_Cont_PT	0.0165	(0.452)	–0.124	(0.461)	–0.0198	(0.459)
External_Cont_AD	0.0154	(1.097)	–0.119	(1.101)	–0.0790	(1.095)
Constant	–22.40***	(1.215)	–22.50***	(1.506)	–22.04***	(1.143)
Border_AD	7		7		7	
Border_BE	7		9		10	
Border_PT	10		10		10	
Border_FR	9		10		10	
Border_DE	8		10		11	
Border_NL	9		12		12	
Border_UK	9		12		12	
Border_IT	9		12		13	
Observations	6,333		6,333		6,305	
R^2	0.890		0.891		0.889	

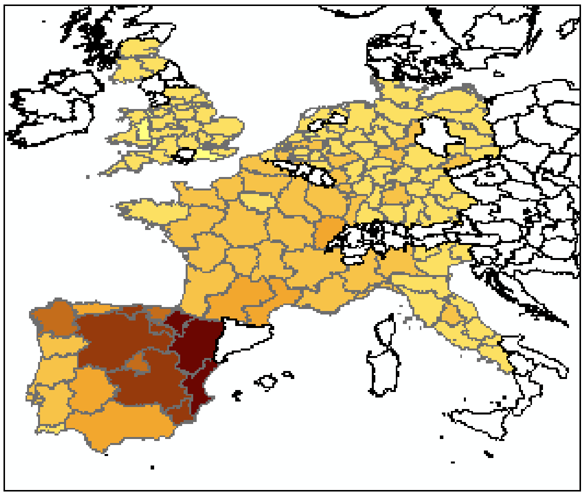
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
All regressions include an "origin by year", "destination by year" and "country destination" fixed effect.
Intraregional trade flows excluded in all models.
Tijt_corr = $\frac{T_{ijt}^{eu}}{Y_{i,t}Y_{j,t}}$

FIGURES

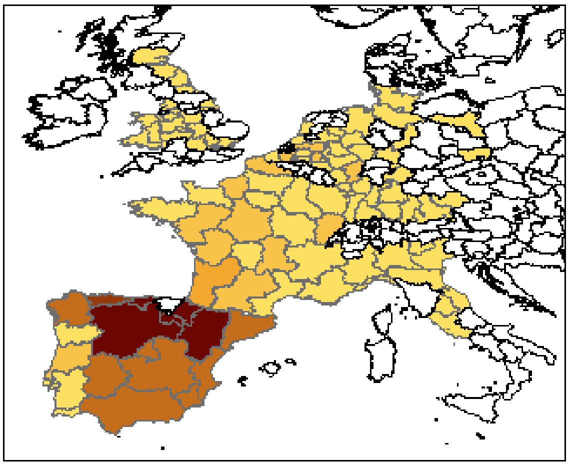
Figure 1: Main Intra-and Inter-National Trade Flows by Road from Selected Regions (€). Average Flows for 2004–2007.



Madrid



Cataluña



País Vasco

Legend

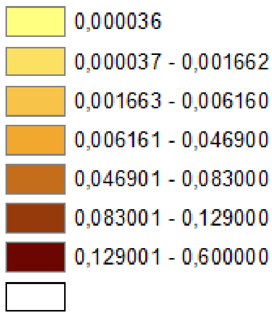
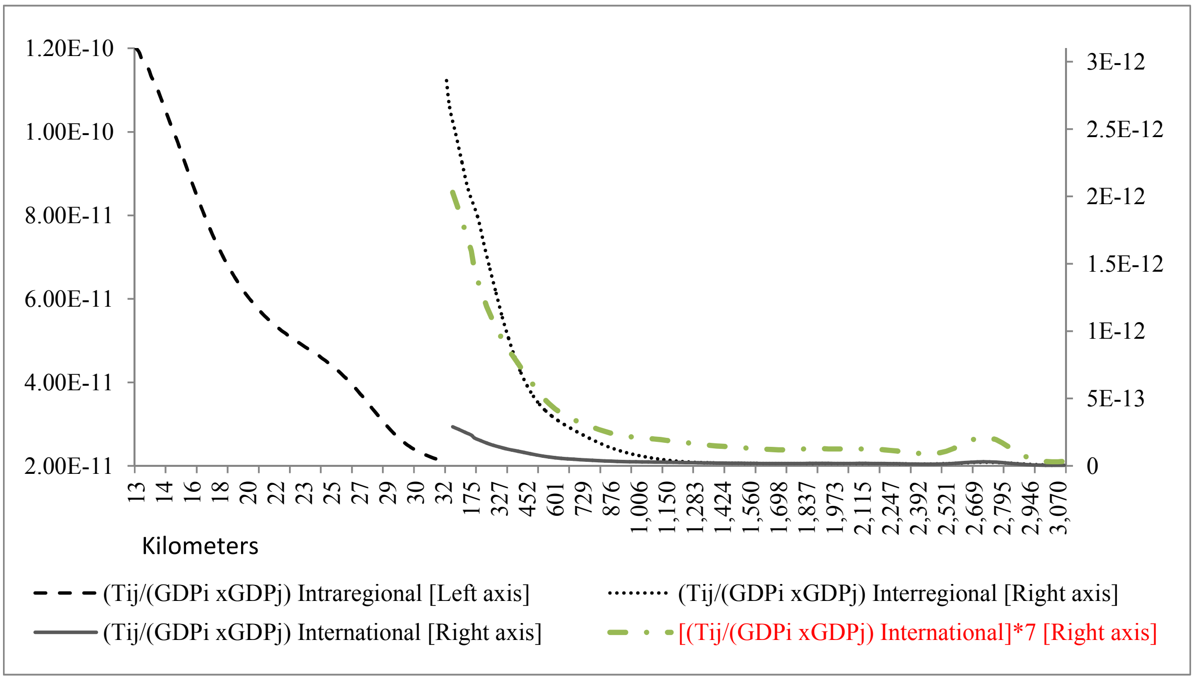


Figure 2: Kernel Regression: Intra- & Inter-National Trade Relative to GDP (NUTS-2 Region-to-Region) on Distance. Zero Flows Excluded. (€). 2004–2007.



APPENDIX

1. Method for Estimating of Region-to-Region Inter-National Trade Flows

We base our method for estimating region-to-region international flows between Spain and eight European countries on a combination of region-to-region freight statistics for Spanish trucking with international price indices for each region-country variety as deduced from official trade data (Gallego and Llano, 2012). Let us start by considering the international export *by road* of products *g* originating in region *i* of country *e* = *Spain* to region *j* of country *u,* $T_{ij}^{eugR} = \left(Q_{ij}^{eugR} P_{ij}^{eugR}\right)$, which can be decomposed into volumes (*Q*) and unit prices (*P*). Superscript *R* emphasizes our exclusive focus on deliveries by road (Spanish trucks). For trade volumes Q_{ij}^{eugR} we rely on the survey published by Spain's Ministry of Public Works and Transport (Ministerio de Fomento) on intra- and inter-national transport flows by road. Like its counterparts in the rest of the EU, this survey provides, for each year *t* and type of product (*G* = *24*; cf. Annex for details), a rich set of variables covering origin and destination flows (in tons) for intra-regional, inter-regional and inter-national shipments. Flows within Spain (both intra- and inter-regional) have been the object of previous analyses (Garmendia et al., 2012; Llano-Verduras et al., 2011; Ghemawhat et al., 2010; Requena and Llano, 2010; Llano et al., 2009). Uniquely in this paper, we use an extended sample drawn from the same survey. This sample specifies the region-to-region dimension of inter-national flows originating in Spanish regions and destined for regions (Nuts-2 level) in Spain's main European partner countries. It is important to remark that the survey covers *volumes* transported by *Spanish* heavy trucks exclusively (F_{ij}^{eugR}); it does not cover any deliveries by international transporters. Insofar as trade volumes are concerned, the dataset is therefore similar to that used in classic papers of the border-effect literature (McCallum, 1995; Helliwell, 1996; Anderson and van Wincoop, 2003; Feenstra, 2002; Hilberry and Hummels, 2008; Helble, 2007). Their dataset also used road transport or truck deliveries as a proxy for trade flows. It is equally important to remark that there is no information on the unit price of products delivered by road, whether domestically or internationally. Consequently, price vectors ($\hat{P}_{ij}^{eugR}$) have had to be estimated on the basis of alternative statistics. For internal flows (intra- and inter-regional), price vectors ($\hat{P}_{ij}^{eugR}$, where e = u = Spain) for each year *t*, product *g* and exporting region *i* were obtained with a large dataset of producer prices for manufactured and agricultural goods (Llano et al., 2009). For international price vectors ($\hat{P}_{ij}^{eugR}$, where $e = Spain$, $u \neq Spain$), we begin (as described in Gallego and Llano, 2012) with a very detailed international trade dataset (*k = 17,000 varieties*) and compute region-country-product specific prices (unit values) for the products considered in the freight statistics (*G = 24 products*). We obtain the *implicit price* for each region-country-product transported by road by means of a weighted average of the *variety price k* included in each product *g*. Each price thus takes into account the *quality* specialization of the internal *variety-mix* within each product *g* ($\{k_1, k_2, \ldots\} \epsilon\ g_1$) for each exporting Spanish region *i* and each importing country *u*. The process also includes a set of filters to control for outliers[17].

2. Testing Dataset Consistency

Here we test our dataset's consistency with official trade statistics. This comparison should be made at a common level of disaggregation for both sources: i.e., with the usual product (*g*)–region (*i*)–country (*u*) breakdown for both volume (tons) and monetary value (euros). **Table I-A** shows our correlation coefficients for the comparison of pairs of equivalent vectors—the official statistics for and our estimate of international export flows. It considers four possible categories of outflows from fifteen Spanish regions (Nuts 2) to the regions of the eight European countries considered: (1) We compare the vectors of aggregate flows measured in volume, for each year *t*, each Spanish region of origin *i* and each country of destination *u* ($\sum_g Q_{ijt}^{eugR};\ \sum_g F_{ijt}^{eugR}$). To distinguish the two available sources, we use the letter *Q* for volume flows drawn from official trade data (in tons) and the letter *F* for corresponding freight flows from our dataset (also in tons). (2) We make the same comparison between the equivalent vector in monetary flow ($\sum_g T_{ijt}^{eugR};\ \sum_g \hat{T}_{ijt}^{eugR}$), where *T* designates official trade data and $\hat{T}$ our estimation (in euros). (3) Finally, the same comparison is repeated once more with the product's specific flow vectors, expressed in both quantity ($Q_{ijt}^{eugR};\ F_{ijt}^{eugR}$) and current euros ($T_{ijt}^{eugR};\ \hat{T}_{ijt}^{eugR}$). In every case, each vector contains bilateral flows for four consecutive years (2004–2007). To make this point clearer, every vector that includes flows for all the years appears with the suffix *t*.

[17] Because of data limitations, our approach assumes that all exports in year *t* of product *g* from a Spanish region *i* to a European country *u* have the same price for all the regions *j* in the importing country. However, we achieve variability in all other dimensions (*t*, *i*, *u*) for the export price of a specific product *g*.

Table I-A. Correlation between Pairs of Vectors on International Trade. Flows in Volume and Monetary units. Region-to-Country level (Nuts 2-0). 2004–2007

Trade (AEAT) vs Freight (EPTMC)	Units	EXPORT
$\sum_g T_{ijt}^{eugR}$; $\sum_g \hat{T}_{ijt}^{eugR}$	Euros	0.8763*
T_{ijt}^{eugR} ; $\hat{T}_{ijt}^{eugR}$	Euros	0.7744*
$\sum_G Q_{ijt}^{eugR}$; $\sum_G F_{ijt}^{eugR}$	Tons	0.9140*
Q_{ijt}^{eugR} ; F_{ijt}^{eugR}	Tons	0.7929*

* Significance at the 1-percent level.
Countries included: AD, BE, DE, FR, IT, NL, PT and UK.
All Spanish regions except: ES53 (Balearic Islands), ES63 (Ceuta), ES64 (Melilla) and ES70 (Canary Islands).

The results in **Table 2** show high levels of correlation between all pairs of vectors for tons and euros. In general, coefficients are statistically significant. The smallest are found in the monetary flows, because of the difficulty of measuring aggregated prices for products (*g)* from disaggregated variety unit-values (*k*). However, the high and significant correlation between coefficients in these cases indicates the aptness of our method for estimating price vectors. This result also illustrates how well our freight-flow dataset functions as a proxy for international trade flows. The high correlation of coefficients in these two datasets allows us to disregard potential contaminations of our dataset with regard to transit flows, multimodal connections and logistical movements that are internal to firms but imply no economic transactions.

ANNEX

Table 4: Countries and Regions Included in the Sample. 2004–2007.

Countries	NUTS 0
Andorra	AD
Belgium	BE
Germany	DE
Spain	ES
France	FR
Italy	IT
Netherlands	NL
Portugal	PT
United Kingdom	UK

Spanish Regions	NUTS 2
Andalucía	ES61
Aragón	ES24
Asturias	ES12
Cantabria	ES13
Castilla y León	ES41
Castilla La Mancha	ES42
Cataluña	ES51
Comunidad Valenciana	ES52
Extremadura	ES43
Galicia	ES11
Comunidad de Madrid	ES30
Región de Murcia	ES62
Navarra	ES22
País Vasco	ES21
La Rioja	ES23

Table 5: Products (G) Included in the Sample. Based on the NST Classification.

Code	Description	Code	Description
1	Grains	16	Natural or manufactured fertilizers
2	Potatoes, other fresh or frozen vegetables, fresh fruit.	17	Coal chemicals, tar
3	Livestock, sugar beets	18	Chemicals other than coal chemicals and tar
4	Wood and cork	19	Pulp and waste
5	Textiles and residuals, other raw materials of animal or vegetable origin	20	Vehicles and transport equipment, machinery, engines (assembled or not) and parts
6	Food and fodder	21	Metalware
7	Oil	22	Glass, glassware, ceramic products
8	Solid mineral fuels	23	Leather, textiles, clothing, miscellaneous manufactured articles
9	Crude oil	24	Various items
10	Petroleum products		
11	Iron ore, scrap, blast-furnace dust		
12	Minerals and non-ferrous residuals		
13	Iron products		
14	Cement, lime, manufactured building materials		
15	Raw or manufactured minerals		

Source: Permanent Survey on Road Transport of Goods, Ministry of Public Works and Transport (Ministerio de Fomento).